International Journal on Artificial Intelligence Tools

About Norms and Causes

DANIEL KAYSER, FARID NOUIOUA
LIPN UMR 7030 du C.N.R.S. Institut Galilée - Univ. Paris-Nord
93430 Villetaneuse, FRANCE
dk, nouiouaf@lipn.univ-paris13.fr

Knowing the norms of a domain is crucial, but there exists no repository of norms. We propose a method to extract them from texts: narrative texts generally do not describe a norm, but rather the discrepancies between what actually happened, and the corresponding normal sequence of events. Answers about the causes of an event often reveal the implicit norms. We apply this idea to the domain of driving, and validate it by designing algorithms that identify, in a text, the "basic" norms to which it refers implicitly.

Keywords: Norms; Nonmonotonic Logic; Causal Reasoning, Natural Language Semantics

1. Introduction

1.1. *Motivation*

Artificial Intelligence is both a Science and a Technique. As a technique, its achievements and its limits are well known and well documented. As a Science, it plays an active role, together with other disciplines belonging to what is now known as cognitive studies , to shed new light on issues that had long been debated: many questions traditionally considered as pertaining to Philosophy, such as the idealization of rational thinking by various kinds of logic, the distinction between word and concept, the notions of meaning, knowledge, and even consciousness, have changed status and are now hot scientific topics.

Among the most original developments due to A.I., restoring to favor the notion of exception in Science should not be played down. The idea of a theory admitting exceptions was long scorned (even by A.I. insiders! see e.g.[3]), because it was allegedly irrefutable: if an experiment confirms the theory, that is fine with it; and if it does not, just add one new exception to the theory and keep it!

Rejecting exceptions forced to elaborate a high standard of demand on scientific theories. This was, and still is, an excellent incentive in 'hard' Sciences, but it may have pushed some human sciences towards dead ends.

In the vast majority of everyday situations, humans do not act as scientists! They have to reach quickly a decision, on the basis of incomplete or uncertain information, and with a partial knowledge of the domain concerned. Achieving such a feat hundreds of times a day is impossible, for complexity reasons, if knowledge is stored and used as a list of truth-preserving rules. To make it possible, knowledge must be structured as simple general rules with exceptions.

A.I. has developed a sub-field, Knowledge Representation, to study a number of techniques that enable computers to reason under the same constraints of incompleteness, and the same requirement of speed as humans do: the best attempts are semantic networks with exceptions[9], non-monotonic logics[1], connectionism[21], inheritance schemes[10]. This endeavor has shown, among other things, that there is a scientific way of handling exceptions: criteria such as parsimony in the coding of knowledge, or algorithmic complexity in the derivation of the most useful results[4] do justify the use of exceptions, without justifying whatever theory you have in mind.

But this (re)discovery of the role of exceptions leads to an interesting engineering challenge: for a given domain, what are the general rules, and how do exceptional situations affect them? Structuring our everyday knowledge in terms of usual and exceptional factors is a point of paramount importance. As the scientific investigation is more accustomed to look for every factor of influence than to elaborate priorities among them, we need to develop a methodology in order to bring out a hierarchy with, at one end, important factors playing a role of full member in a rule, and at the other end, subsidiary factors to be considered only in exceptional cases. This is the reason why eliciting the norms of a domain is important.

1.2. *Plan of the paper*

This paper presents a methodology to extract norms from texts concerning a specific domain. We cannot expect to do so from an automatic examination of the texts: as every reader knows them, the norms are never spelled out. However, it has often been noticed that many texts, mostly narratives, describe the discrepancies between what actually happened, and the corresponding normal sequence of events. This is of course not enough to infer the norm from the text. After a short discussion on the notion of norm (section 2), we introduce in section 3 the idea consisting in using causation as leverage to find the norms.

Section 4 describes the method. Section 5 presents a reified first-order logic to cope with the problem of time and modalities. Section 6 illustrates the whole process with examples.

2. About Norms

2.1. *Polysemy of " norm"*

The word "norm" refers to at least two different ideas: one is *prescriptive*; the adjective corresponding to it in this case is "normative"; the other is *descriptive*, and the corresponding adjective is "normal". The majority of A.I. works about norms use the first sense of the word[2,8] and concern various kinds of deontic logic[16]. However, the interest that we have in norms is motivated by the second sense: we want to describe what is said to be normal.

Notice that other (intermediate?) senses also exist: for instance, a norm can be an idealization of a practice, without being a prescription; in this sense, formal logic can be looked at as the norm of human reasoning: it is not descriptive; calling it a norm is not claiming that normal humans follow logical patterns of reasoning; but it is not truly prescriptive either: not following these patterns is not considered as ill mannered, and it has no judicial consequences!

2.2. *Knowing what is normal*

The importance of knowing what is normal emerged progressively in various subfields of A.I.:

(1) knowing what will normally happen after an action is vital for predicting and planning;
(2) in the reverse direction, this knowledge is necessary to diagnose which action may have caused a given state-of-affairs;
(3) it has often been pointed out that text understanding presupposes from the reader a knowledge of the norms; hence, it is very rare to find a text spelling out the norms, and it is much more frequent to describe a situation or an unfolding of events by its deviations from the underlying implicit norm.

Point 3., the needs related to text understanding, has led early A.I. researchers like[5] to focus on the representation of normal uses of objects and normal courses of events, and the first proposed tools to solve this kind of problems were *frames*[17] and *scripts*[22]. Capturing the ability to derive conclusions that normally follow from premises has later been the major incentive for developing, in the 1980's, non-monotonic logics. There is no need to remind the importance for A.I. of the developments of this topic.

2.3. *Truth-based vs. Norm-based inferences in Natural Language*

Traditional NL semantics focuses on truth conditions. A text is said to entail the propositions that hold true whenever the conditions making the text true are fulfilled. This is a very weak notion of entailment. Consider for example:

(R) *I was driving on Main Street when a truck before me braked suddenly.*

The ***truth-based*** entailments of (R) contain propositions like: "there exists a time t such that I was driving at t, and there was a truck T before me at t, and T braked at t."

The ***norm-based*** entailments contain e.g.: "Both T and me were moving in the same direction with no other vehicle in between. The distance between us was less than 100 meters."

Of course, the propositions of the latter list might be false while those of the former list are necessarily true if (R) is correct. Nonetheless:

- every reader of (R) will take them, at least provisionally, as legitimate inferences,

- the author of (R) knows that explicit indications should be provided somewhere in the text, in order to block these inferences, if they are incorrect.

Norm-based inferences are thus as rightful and much richer than truth-based ones. According to the usual tripartition between syntax, semantics and pragmatics, they might be classified under pragmatics rather than under semantics, but no matter the label, they are extremely important. The problem is how to find the norms enabling to derive them.

2.4. *How to find the norms?*

There is no repository of norms ruling a given domain (norm is taken here in the sense corresponding to 'normal', not 'normative'). A somehow similar problem arises in other areas of Knowledge Representation. In both cases, there exist good frameworks (non-monotonic logics, description logics, and so on), but filling the framework with actual data is a difficult task (e.g. the CYC project[13]).

The data one can start with, in both cases, is the huge amount of textual data now available under electronic form. But the problem is easier when the goal is to elicit so-called ontologies from such texts. Kind-of and part-of hierarchies, which are the critical ingredients of an ontology, are more or less explicitly present in dictionaries, thesauri, glossaries, etc. There is no such starting point in the case of norms.

3. About Causes

3.1. *Is there a definition of cause?*

The notion of cause is deeply rooted in human cognition, and there is an abundant literature witnessing the fact that it has been a focus of interest at all times. Yet, strangely enough, there is very little consensus about the nature of causation. Some philosophers consider it to be a "real"feature of the world[11]: according to this point of view, A intrinsically causes B, no matter what humans think about it; others consider it as a psychological notion invented by humans to make sense of their observations of the world, the world in itself having nothing causal (this was the point of view developed for instance by Bertrand Russell; see discussion e.g. in[19]).

Our discipline, A.I., has not to take side in the metaphysical issue as to whether causes and effects "really" exist. But it has to give an account of an ubiquitous kind of reasoning: *causal reasoning*, because it is uncontroversially very useful: building systems for planning, diagnosing or predicting, i.e. for the most useful A.I. applications, is absolutely impossible without reasoning on causes and effects.

In the kind of applications that A.I. has to care for, what is worth being considered as a cause must be something we can act upon. Gravitation may well be a cause of a glass being broken; it leads nowhere to blame gravitation for the damage that happened; finding who struck the glass is a more useful move! But filtering the potentially endless list of factors that can be considered as having caused a given fact, by requiring that we must be able to act on them, is not enough. The list remains too large to be of practical use. Now, in a given context, only few causes come to the mind. Mackie's well-known example[14] of an explosion happening when a cigarette is lit is revealing: if someone is asked for the cause and knows that the event took place at home, a sensible answer can be: "*perhaps, the gas was leaking*"; if it took place in an oil refinery, the answer should be: "*smoking in such places causes explosions*".

3.2. *Relation between cause and norm*

The above example shows that, even if it does not seem so at first sight, the notions of cause and norm are tightly related. As a matter of fact, the reason why the answer differs at home and in a refinery, is that the norms ruling the behavior at home and in a refinery are different. More generally, it seems that the best answer to a *causal* question concerning abnormal events ("*why did this happen*"), is to point to a violated *norm*.

Hence one of the main ideas developed in this paper, is that a good methodology to elicit the norms of a given domain consists in analyzing the answers to questions concerning the cause of abnormal events.

If we are told that, after skidding on bits of gravel in a bend of a highway, a motorcycle went off the road, the reasons that come to mind to explain the accident are the presumably excessive speed of the motorcyclist in the bend, and the unexpected presence of gravel on the road, i.e. two abnormal factors on which one can rather easily act. No one would call a cause of the accident the presence of gas in the tank of the motorcycle, because this is normal, nor would one blame the centrifugal force, because no action can prevent it to exert its effect.

Many other examples (see below) confirm the tendency to ascribe the cause of an abnormal event, e.g. an accident, to an abnormal factor on which a voluntary action could in principle remedy. Therefore, the capacity to find the same causes as a human reader is a good criterion to check whether the set of norms of a given domain is mastered.

4. Method

4.1. *The corpus*

The domain we selected concerns car accidents, because:

- we have already studied, for different reasons, a corpus of car crash reports written by drivers after an accident[23]; the reports are short texts, syntactically simple, and (unfortunately) we can get as many of them as we wish from French insurance companies;
- they describe abnormal events occurring in a domain that is non-trivial, but limited well enough to make plausible an enumeration of all of its norms;

- hey are good representatives of texts requiring the reader to perform norm-based inferences in order to understand what happened.

These texts have however a disadvantage: they are often biased. The author tends to describe the accident from the point of view that minimizes his/her responsibility. But this is not a real drawback: the norms that are important are those which the drivers have in mind, not the set of norms that govern the real events, and these norms are revealed by biased descriptions as well. Anyway no corpus can pretend to contain unbiased descriptions of events.

4.2. *General idea*

In order to study the relationships between norms and causes within the limited domain of car-crash reports, we gathered a sample of these reports and determined, in each case, what we considered to be the cause of the accident. The problem then consists in designing algorithms that copy this behavior. If the algorithm also correctly identifies the cause of the accident in reports taken from outside the sample, this will mean that the set of norms collected is reasonably complete, and it will validate our approach.

This is a hard problem, involving linguistic and knowledge representation issues, as well as reasoning mechanisms.

Linguistic issues are so difficult that, by themselves, they would justify a full project. Some of the difficulties have already been identified in a previous work on the same corpus[23]. Nonetheless, the risk is high of being stuck in nearly insoluble linguistic problems, without ever knowing whether their resolution matters for our purpose. Therefore, instead of going "forward" from the text to the norms, we started "backwards": we assumed the earlier stages to be solved, and concentrate on the last ones. This method secures that we focus only on issues that are necessary.

4.3. *Basic and Derived Anomalies*

The violation of the well-known norm:

(N1) "Under all circumstances, one must have control over one's vehicle."

can explain almost every accident. However, in a text like:

In order to avoid a child suddenly rushing on the street, I swerved and bumped into a parked vehicle.

identifying the cause of the accident as a mere "*loss of control*" would be misleading. As a matter of fact, the violation of (N1) is here a consequence of a more imperative norm:

(N2) "Harming a person must be avoided."

Accordingly, "*because the driver swerved to avoid the child*" is a better explanation of the accident. As accidents are anomalies, we want to trace back their causes to an anomaly; obeying (N2) being a normal behavior, we further identify the norm:

(N3) "Persons must not rush on causeways."

and prefer to ascribe the cause of the accident to the violation of (N3). Notice that this is not the only possible cause: in a densely populated area, the norm is to drive slowly, and it could be argued that the violation of this norm is a better explanation. In either case, the tendency is confirmed: we take as a cause the deliberate violation of a norm, when such a norm comes to our mind.

We sorted out, for each car-crash report of the corpus, what seemed to be the "basic" cause of the accident. This work is delicate. The aim is to extract what a normal reader understands from the circumstances related, but, as we said, the description is often biased, and a normal reader knows that it can be so; therefore, understanding the report does not entail to take every statement of it for the pure truth. Sometimes, the writer's argumentation is so obviously a purely rhetorical game, that the reader is implicitly driven to understand it as an admission of responsibility. A difficult part of the work is thus to determine up to which point the meaning intended by the writer deviates from the literal meaning of his/her text.

Anyhow, the set of causes considered basic, after all the texts have been manually processed, reveals a rather comprehensive subset of the norms of the domain.

4.4. *Layers*

Our problem is to handle the transition from the hundreds of relevant propositions found in the texts, towards an expected small number of norms. It would be foolish to attempt solving it in one single step. We therefore split the problem in layers: the most external one is the mere result of a parser, filtered from the irrelevant elements of the reports. The most internal one, layer 1 also called the "kernel", consists in a very small number of predicates (see Appendix).

Hypothesis H1: All the "basic" anomalies can be explained by means of the predicates of the kernel.

Furthermore, we assume that the representation of the text at layer n can be obtained from its representation at layers $n' > n$ by means of a limited number of inference rules, each one factoring out the common features that govern a specific concern. The whole process is thus a stepwise convergence from a vast variety of situations towards a smaller number of cases.

We consider two groups of layers, roughly described as the linguistic group and the conceptual group.

The *linguistic* group starts from the result of the parser, i.e. a set of syntactic predicates expressing the grammatical relations between words. It translates this set, through several steps, into a set of predications involving a pre-defined list of *semantic predicates*. One step consists in using equivalence tables to replace by a single representative a variety of words having roughly the same meaning in the context of road accidents; another step concerns the resolution of anaphoric co-references; determining the semantic relationships between the elements of a sentence: they do not result immediately from the syntax.

For example, in ***le véhicule A est venu heurter le véhicule B*** (lit. *vehicle A has come to hit vehicle B*), A is the grammatical subject of *come*; B is the object of *hit*. Now, the sentence actually means that vehicle A hit vehicle B, i.e. vehicles A and B are semantically connected as being the respective subject and object of the same verb *to hit*; the verb *to come* is only used here as a "support".

At the end of this process, the text is transformed into a conjunction of literals expressed with predicates taken from a list of around 60 semantic predicates. The next stages use the *conceptual* group of layers, and this process is developed in the rest of the paper.

As said above, the goal is to convert progressively this conjunction into a conjunction of literals of the kernel, from which, according to (H1), the basic anomaly, i.e. the answer to the question "why did the accident happen", can be discovered.

5. Time, modalities, and inference

5.1. *Different times involved*

An important open issue is the connection between the grammatical tenses, found in the text, and the phenomenal time[6]. But in order to find a satisfactory solution, we must first determine precisely what time needs be represented, as several times play a role in our problem:

(i) the (linear) time of the reader: propositions in the text are ordered, but this order does not necessarily reflect the sequence of the events accounted for;
(ii) the (linear) time of the events;
(iii) he (branching) time considered by the agents: each agent indeed knows that only one future will come true, but his/her actions are explainable only by considering the possibility of several of them, among which (s)he tries to eliminate the undesirable ones.

Hypothesis H2: Our goal requires only the explicit representation of (ii).

H2 is a rather strong hypothesis, since our texts abound with lexical items like "*avoid*", "*prepare*", "*expect*" (and their negation) that evoke unrealized futures. However, the author generally makes use of them for argumentation purposes, which provide no significant help in finding the causes of the accident.

Notice that the hypothesis does not consider the unrealized futures as of no import, but only that these futures do not need explicit representation.

5.2. *About anomalies*

Hypothesis H3: The *basic* anomalies are of two kinds; either an agent had to do some action a, had the ability to do a, and did not a; or an external factor that could not reasonably be foreseen explains the accident. The *derived* anomalies are those where an agent should have done an action a, but, due to a basic anomaly, was not in position to do it.

According to *H3*, we need to reason on propositions of the form: MUST-DO a and ABLE-TO-DO a. They look like modals, and indeed they are; the former is clearly a kind of necessity, and the latter, a kind of possibility, but they do not obey the usual duality relationship. For one thing, an agent must not do ¬a every time (s)he is not able to do a. We develop this issue in more details in the next paragraph.

5.3. *States and accessibility between states*

A modal account of the duties and abilities of agents can be given by means of a possible world semantics. The accessibility between states has clearly a temporal flavor. Yet, representing every time point of a sequence of events is unnecessary; in fact, the reports look like a sequence of pictures, rather than like a film, but this metaphor is not fully adequate, since the "pictures" may use predicates that are dynamic in nature. So, a state is not characterized by the propositions which are true at a given time point, but rather by those remaining true during a given interval.

The problem is to determine the intervals. Basically, the idea is to consider an interval as a kripkean world satisfying a predicate of action as well as its relevant effects (ramifications); the accessibility from a world w to another one w' is the result of one of the two following possibilities:

- an agent has decided in w to perform an action; in that case, there exists accessible worlds w' containing its possible outcomes; for example, if the agent decides to brake in w, then w' contains the braking action and either the fact that the speed is low, or that the vehicle stopped. The corresponding transitions are called C-accessibilities;
- an external event, or an involuntary action happened in w; for instance, driving on an icy patch in w leads to a world w' where the control of the vehicle is lost; these transitions are called E-accessibilities.

A subset of the C-accessibilities, named the D-accessibilities, gathers the transitions where the agents perform the actions dictated by their duties. The modality $MUST-DO$ is thus the necessity modality associated, in a kripkean sense, to the D-accessibilities, while the modality $ABLE-TO-DO$ is the possibility modality associated with the C-accessibilities. This explains why they are not dual to each other in the usual sense.

A world, or state, corresponds to an interval in which no change in modality occurs, i.e. what the agent $MUST-DO$ and is $ABLE-TO-DO$ remains the same. On the contrary, the decision of an agent to do (or not to do) an action takes place in a state that strictly precedes the state where the action is performed.

It ensues that anomalies are found in transitions, not in states. As the usual syntax of modal logics is not well adapted to assert statements about transitions, we find more convenient to represent modalities as first-order predicates in a reified logic. Briefly stated, the reification of a logical language L is a language RL where formulas such as $P(x,\ y)$ in L are written $HOLDS(P,\ x,\ y)$ in RL. This has the advantages of allowing to quantify over the predicate names without using higher-order logic, and to consider ordinary statements, of the form HOLDS(predicate-name, agent-name, state), on a par with modal statements, e.g. $MUST-DO$(predicate-name, agent-name, state). Of course, the expressive power of a reified language remains within the limits of expressiveness of first-order logic, so it is impossible to quantify over the set of all predicates; but it is possible to consider some predicates as being objects (hence the name, reification); these objects form a subset of the domain, so one can quantify over this subset, and this is what is needed in practical applications. There is a price to pay: to represent the negation of P, we have to introduce a constant $not-P$ and to explicit obvious facts:

$$(NEG) \quad (\forall\ P,\ x,\ y)\ HOLDS\ (P,\ x,\ y)\ \leftrightarrow\ \neg HOLDS\ (not-P,\ x,\ y)$$

which are given for free in usual logic. Practically, this constraint is not very cumbersome. It would be more tedious if we had to reason on conjunctions or disjunctions, e.g. $HOLDS(P-and-Q,\ x,\ y)$. We never met such needs.

For instance, the two formats of $H3$ are expressed by the formulas (p is the name of a predicate, Ag, the name of an agent, B-An stands for Basic-Anomaly):

$$(F) \quad (\forall\ p,\ Ag,\ t)\ MUST-DO\ (p,\ Ag,\ t)\ \wedge\ ABLE-TO-DO\ (p,\ Ag,\ t)\ \wedge \\ \neg HOLDS\ (p,\ Ag,\ t+1)\ \rightarrow\ B-An$$

$$(F') \quad (\forall\ p,\ Ag,\ t)\ ABNORMAL-PERTURBATION\ (p,\ Ag,\ t)\ \rightarrow\ B-An$$

Whereas in most cases several accessible (future) states are meaningful, $H2$ says that only members of a totally ordered sequence need actually be present to reveal the basic anomalies. States are thus represented as integers, yielding a simplified version of the notion of chronicle [15].

In order to represent a scriptal unfolding of events, we introduce a third modality, NORMALLY. It corresponds to a necessity, in the kripkean sense, with regard to a subset N of the union of C and E-accessibilities. A state is N-accessible if it corresponds to an expected outcome of the current situation (for example, in C: when a driver arrives at a stop sign, s/he is expected to stop; in E: if a traffic light is red, it is expected to turn green). As agents are expected to comply with their duties, D is a subset of $N\ \cap\ C$.

In our language, this modality is expressed by a pseudo-modal predicate: $NORMALLY\ (p,\ Ag,\ t)$. The fact that D is a subset of N is thus written:

$$(D) \quad (\forall\ p,\ Ag,\ t)\ MUST-DO\ (p,\ Ag,\ t)\ \rightarrow\ NORMALLY\ (p,\ Ag,\ t)$$

5.4. *Inference rules*

Introducing "normal" transitions naturally leads to using non-monotonic inference rules. We use a fragment of Reiter's default rules [20]. This choice is motivated by reasons of clarity, but as defaults easily translate into auto-epistemic logic [7,12], we can take advantage of several existing deductive systems. We write $A \ : \ B$ and $A \ : \ B[C]$ as shorthands for respectively the normal default $\frac{A \ : \ B}{B}$ and the semi-normal default $\frac{A \ : \ (B \ \wedge \ C)}{B}$.

The basic default rule is:

$$(N) \quad NORMALLY \ (p, \ Ag, \ t) \ : \ HOLDS \ (p, \ Ag, \ t+1)$$

i.e. the transitions $(t, t+1)$ in the actual unfolding of states are normal ones. Of course, since our texts report on accidents, there must be at least one exception, i.e. one actual transition that is abnormal.

In order to avoid an uncontrolled proliferation in the number of extensions, we appeal to defaults only when we have strong reasons to believe that a report could imply an exception to the rule we are expressing. In all other cases, even when exceptions are conceivable, we use material implications.

As an example, consider the following rule, which captures an abductive reasoning:

$$HOLDS \ (Same\_File, \ Ag, \ Ag', \ t) \ \wedge \ HOLDS \ (Crash, \ Ag, \ Ag', \ t) \ : \\ HOLDS \ (Is\_follower, \ Ag', \ Ag, \ t-1)$$

(if Ag' bumps into Ag while both of them are in the same file, it is most likely that Ag' was the follower of Ag in that file). We use a default here because, although we do not have the case in our sample, it is not implausible that Ag' hit Ag at the end of overtaking Ag's follower; we hypothesize that in this case, the report would give enough elements to derive $\neg HOLDS \ (Is\_follower, \ Ag', \ Ag, \ t-1)$, thereby preventing the application of the rule.

The reader probably notices that $HOLDS$ gets a varying arity (3 or 4). This is clearly forbidden in first-order logic. We present it that way for clarity [actually $HOLDS$ is ternary and, when needed, a binary function combines together the extra-arguments. The actual expression is thus: $HOLDS \ (COMBINE \ (Same\_File, \ Ag), \ Ag', \ t)$]. The same trick is used for the pseudo-modal predicates.

The inference engine uses both "strict" rules (like (D) above) and "default" rules (like (N)); it stops as soon as one of the rules (F) or (F') produces the atom $B - An$ (basic-anomaly).

6. Examples

Space limitations prevent us from showing but very simple examples. A significant minority of reports have this level of simplicity, but a majority of them are far more complex.

6.1. *Text B21*

Our first example is text B21 of our corpus:

Nous nous sommes arrêtés pour laisser passer un véhicule de pompiers, la voiture qui nous suivait nous a alors heurtés.

We stopped to let a vehicle of firemen through; the car following us then bumped on us.

Assuming as said above, that the linguistic issues are correctly handled, we start with three states 0, 1 and 2; the initial state 0 is by default a state where all vehicles are under control.

State # 1.

$HOLDS\ (Stopped,\ A,\ 1)$; the reason given to this event, the firemen, is purely contextual: we thus omit it.

State # 2.

$$HOLDS\ (Crash,\ A,\ B,\ 2) \qquad HOLDS(Is\_follower,\ B,\ A,\ 2).$$

Inferences

Some predicates are declared static, and are endowed with forward default persistence, i.e.

$$(S) \quad STATIC\ (p)\ \wedge\ HOLDS\ (p,\ Ag,\ t)\ :\ HOLDS\ (p,\ Ag,\ t+1)$$

This assumption is usual[15]. Here, and in several other texts, we need also a kind of abductive reasoning entailing backward persistence. Being static is not enough for being backward persistent, so we declare which predicates have this feature on a case-by-case basis. Here, we do have:

$$HOLDS\ (Is\_follower,\ Ag,\ Ag',\ t)\ :\ HOLDS\ (Is\_follower,\ Ag,\ Ag',\ t-1)$$

his default yields: $HOLDS\ (Is\_follower,\ B,\ A,\ 1)$. Another rule (a strict one) tells:

$$(Cr) \quad (\forall\, Ag,\ Ag',\ t)\ HOLDS\ (Crash,\ Ag,\ Ag',\ t)\ \rightarrow\ \neg HOLDS\ (Stopped,\ Ag',\ t)$$

(for an agent Ag' to bump into Ag at time t requires that Ag' is not stopped at time t).

We thus get: $\neg HOLDS(Stopped, B, 2)$.

Norm $(N1)$ (see {}4.3.) translates into the rule:

$$(\forall\ Ag,\ t)\ MUST-DO\ (Control,\ Ag,\ t)$$

Therefore, $MUST - DO\ (Control,\ B,\ 0)$ from which (D) and (N), with $t\ =\ 0$, give: $HOLDS\ (Control,\ B,\ 1)$.

We also have:

$$HOLDS\ (Is\_follower,\ Ag',\ Ag,\ t)\ \wedge\ HOLDS\ (Stopped,\ Ag,\ t)\ :\ MUST - DO\ (Stopped,\ Ag',\ t)\ [HOLDS\ (Control,\ Ag',\ t)]$$

which means that if Ag' follows Ag, and Ag stops, then Ag' must stop too, unless Ag' is not under control.

This rule provides: $MUST - DO\ (Stopped,\ B,\ 1)$. Finally,

$$(\forall\ Ag,\ t)\ HOLDS\ (Control,\ Ag,\ t)\ \rightarrow\ ABLE - TO - DO\ (Stopped,\ Ag,\ t)$$

i.e. for a vehicle, being under control implies being able to stop. This rule gives us: $ABLE - TO - DO\ (Stopped,\ B,\ 1)$ which completes the premises of (F) (5.3.) with $p\ =\ Stopped,\ Ag\ =\ B,\ t\ =\ 1$, and allows to derive $B - An$.

This derivation stops the process. We can answer the question "Why did the accident happen?" and the answer, provided a simple NL generator is written, is: "because vehicle B did not stop at state 2".

6.2. *Text B7*

Les deux véhicules étaient en stationnement perpendiculairement. En l'absence des conducteurs, le frein du véhicule B a lâché. Entraîné par la pente, le véhicule B a traversé la chaussée et a heurté le véhicule A.

The two vehicles were parked perpendicularly. In the absence of the drivers, the brake of vehicle B released. Because of the slope, vehicle B crossed the street and bumped into vehicle A.

Here, the text is decomposed in three states:

State # 0.

The default state 0 does not apply here, since two vehicles are parked and the drivers are absent. We have:

$$HOLDS\ (Parked,\ A,\ 0) \qquad HOLDS\ (Parked,\ B,\ 0)$$

$$\neg HOLDS\ (With\_Driver,\ A,\ 0) \qquad \neg HOLDS\ (With\_Driver,\ B,\ 0)$$

State # 1.

The brake of vehicle B releases; this event is considered punctual, and the state contains the consequence of the event, i.e. vehicle B is no more at rest.

$$HOLDS\ (Abnormal\_Brake,\ B,\ 1) \qquad \neg HOLDS\ (Stopped,\ B,\ 1)$$

State # 2.

This is the state where B hits A.

$$HOLDS\ (Crash,\ A,\ B,\ 2)$$

Inferences

We first apply the rule:

$$(\forall\ Ag,\ t)\ HOLDS\ (Parked,\ Ag,\ t)\ \rightarrow\ HOLDS\ (Stopped,\ Ag,\ t)$$

(A parked vehicle is stopped), and obtain:

$$HOLDS\ (Stopped,\ A,\ 0)\quad HOLDS\ (Stopped,\ B,\ 0).$$

Parked is forward-persistent, i.e. we have $STATIC\ (Parked)$, hence through (S):

$$HOLDS\ (Parked,\ A,\ 1)\qquad HOLDS\ (Parked,\ A,\ 2).$$

$$HOLDS\ (Stopped,\ A,\ 1)\qquad HOLDS\ (Stopped,\ A,\ 2).$$

The default (S) is blocked for B, as we have:

$$\neg HOLDS\ (Stopped,\ B,\ 1)\ \ \textit{and by contraposition}\ \ \neg HOLDS\ (Parked,\ B,\ 1)$$

Moving (i.e. not being stopped) is also forward persistent: $STATIC\ (not-Stopped)$. From (NEG) and $\neg HOLDS\ (Stopped,\ B,\ 1)$ we derive $HOLDS\ (not-Stopped,\ B,\ 1)$ and through (S) and (NEG) again: $\neg HOLDS\ (Stopped,\ B,\ 2)$

Incidentally, the same literal is produced by another way: the rule (Cr) of example 6.1.

We apply now the following rule:

$$(\forall\ Ag,\ t)\ HOLDS\ (Abnormal\_Brake,\ Ag,\ t)\ \wedge\ \neg HOLDS\ (Stopped,\ Ag,\ t)\ \rightarrow \\ \neg HOLDS\ (Control,\ Ag,\ t)\ \wedge\ \neg ABLE-TO-DO\ (Control,\ Ag,\ t-1)$$

(if the brake of a vehicle is abnormal at a state t where the vehicle is moving - the vehicle may have been at rest or not at state $t-1$-, then the vehicle is no longer controlled at state t and the driver was not in position at state $t-1$ to keep control over it, because the problems with the brakes are considered to be unpredictable).

This rule yields here:

$$\neg HOLDS\ (Control,\ B,\ 1)\qquad \neg ABLE-TO-DO\ (Control,\ B,\ 0).$$

This last fact prevents (F) from firing: as B was not in position to control, his lack of control is not the basic anomaly. As a matter of fact, we are in presence of what we called (5.2.) a derived anomaly. But we also have the rule:

$$(\forall\ Ag,\ t)\ HOLDS\ (Abnormal\_Brake,\ Ag,\ t)\ \wedge\ \neg HOLDS\ (Stopped,\ Ag,\ t)\ \rightarrow \\ ABNORMAL-PERTURBATION\ (Brake,\ Ag,\ t)$$

that expresses the fact that an abnormal brake in a moving vehicle is to be considered as a major perturbation. We thus have: $ABNORMAL - PERTURBATION$ $(Brake,\ B,\ 1)$ and, through (F'), we derive the basic anomaly. The cause of the accident is thus taken here to be an external, unpredictable, event: the brake of vehicle B had an anomaly.

Several other derived anomalies are also discovered; for instance we have a rule expressing that if a collision happened between two vehicles, each of them constitutes an obstacle for the other. Moreover, except under some specific conditions, when a vehicle has an obstacle, it must stop. This allows to derive $MUST - DO$ $(Stopped,\ B,\ 1)$; as we derived (by two different paths) $\neg HOLDS$ $(Stopped,\ B,\ 2)$, (F) has interesting premises for detecting a basic anomaly; but, since B is not under control at state 1, we also have $\neg ABLE - TO - DO$ $(Stopped,\ B,\ 1)$. Therefore (F) cannot fire and this is only a derived anomaly.

7. Perspectives and conclusion

7.1. *Current state and short-term perspective*

We have analyzed by hand a set of 73 reports. In each case, we have identified the basic anomaly and the sequence of states that corresponds to the unfolding described by the report.

We have written 74 rules and defaults to handle what we call layers 1 and 2. The linguistic group of layers is well advanced, and most rules of layer 3 are already written too. Analyzing new texts will certainly show the need for new rules, or for generalizing the existing ones, but we are fairly confident that the size of the whole enterprise remains manageable: at worst, a few hundreds of rules should be necessary. We intend to complete shortly the other layers, and to validate the approach by testing them on fresh reports.

Several taggers and parsers are available for French, we tested a tagger and the results were quite satisfactory; the specificities of the reports -as well as the state of the art for general-purpose parsers- are such that the parses we obtain are less suitable. We therefore developed our own parser, which fares better on the corpus, but its scaling up to the new reports remains to be assessed.

he derivations are currently performed by hand. There exist a number of deduction engines working with various subsets of non-monotonic logics. We plan to switch in the near future from manual to automatic deductions.

If, at this stage, everything works correctly, we will consider that the norms of the domain are suitably represented by the default rules and by some strict rules, since as we said, we appeal to defaults not if exceptions are conceivable, but only when exceptional cases are likely to be present in a report. This set is obtained entirely from our intuition, and the only help provided by the computer is to check, a posteriori, that it is reasonably complete; it is the case if the computer derives from a report the same cause of the accident as a human reader.

7.2. *Longer term perspective and conclusion*

As a further step of the study, it would be interesting to propose automatically one or more new norms, or a generalization of existing ones, when the cause found by the computer does not coincide with the cause found by the human reader. Learning algorithms could be adapted in order to help the elicitation of norms. But the validation of the norms by an expert would remain compulsory, so at the end, we would have a semi-automatic method to extract norms from texts.

But the important issue is not only the success or failure of getting this work done more or less automatically. We already have good reasons to believe that the domain of road accident is explainable by a limited number of norms, and that a relatively small number of rules is enough to find which ones among them are violated. We will thus try to apply a similar methodology to other domains, by asking experts what they perceive to be causes of anomalies, since this seems to be a good way to elicit the norms.

t may turn out that the number of norms in various fields is not as formidable as one could think. And representing the norms is of paramount importance to extend the inference capabilities far beyond what is warranted by truth-conditional semantics, so our approach could help in lifting a major bottleneck for A.I. Another possibly fruitful idea would be to classify texts by the norms they are referring to, because it might open interesting tracks for indexing documents.

Acknowledgments

This work is being done in the framework of the franco-algerian project "modèles d'action" (no. 02 MDU 542) funded by the Comité Mixte d'Évaluation et de Prospective (C.M.E.P.). The authors are indebted to Françoise Gayral, François Lévy, and Adeline Nazarenko for helpful remarks on a previous version of this text.

Appendix A. The "Kernel"

Layer 1 contains 7 reified predicates. 5 of them have a pair ¡vehicle, state¿ as arguments: Stopped, Starts, Runs_slowly, Runs_backwards, Control; all of them express that the vehicle satisfies the named property during the time interval corresponding to the given state. The last two predicates are: Changes_speed (first argument is '+' or '-' depending on whether the driver speeds up or brakes, the two remaining arguments as before) and Abnormal-Perturbation (name of the disruptive factor, vehicle, state). The whole set of predicates and rules of layers 1 and 2 can be found in [18].

References

1. AI 1980. Special issue on non-monotonic logic. Artificial Intelligence 13 (1-2).
2. Boman, M. 1999. Norms in Artificial Decision-Making. Artificial Intelligence and Law, 7(1): 17-35.

3. Brachman, R.J. 1985. I Lied about the Trees (or Defaults and Definitions in Knowledge Representation). A.I. Magazine 6(3): 80-93, Fall.
4. Cadoli, M.; Donini, F.M.; Schaerf, M. 1996. Is intractability of nonmonotonic reasoning a real drawback? Artificial Intelligence 88(1-2): 215-251.
5. Charniak, E. 1973. Jack and Janet in search of a theory of knowledge. Proc. 3rd Int. Joint Conf. on AI 337-355, Stanford, August.
6. De Glas, M.; Desclés, J.P. 1996. Du temps linguistique comme idéalisation d'un temps phénoménal. Intellectica 23: 159-192.
7. Denecker, M.; Marek, V.W.; Truczynski, M. 2003. Uniform semantic treatment of default and autoepistemic logics. Artificial Intelligence 143(1): 79-122. 401–413.
8. Dignum, F.; Kinny, D.; and Sonenberg, L. 2002. From Desires, Obligations and Norms to Goals. Cognitive Science Quarterly 2(3-4): 405-427.
9. Fahlman, S.E. 1979. NETL - A System for Representing and Using Real-World Knowledge. M.I.T. Press.
10. Horty, J.F.; Thomason, R. H.; Touretzky, D. S. 1990. A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks. Artificial Intelligence 42(3): 311-348.
11. Kistler, M. 1999. Causalité et lois de la nature Paris: Vrin, coll.Mathesis.
12. Konolige, K. 1988. On the relationship between Default Logic and Autoepistemic Logic. Artificial Intelligence 35(3): 343-382.
13. Lenat, D.; Guha, R. 1990. Building Large Knowledge-Based Systems. Addison-Wesley.
14. Mackie, J.L. 1974. The Cement of the Universe: A Study of Causation. Oxford University Press.
15. McDermott, D.V. 1982. A Temporal Logic for Reasoning about Processes and Plans Cognitive Science 6: 101-155.
16. McNamara, P.; Prakken, H. eds. 1999. Norms, Logics and Information Systems. New Studies in Deontic Logic and Computer Science. Vol.49 Frontiers in Artificial Intelligence and Applications, IOS Press.
17. Minsky, M. 1974. A Framework for Representing Knowledge. A.I.Memo 306, M.I.T. (reprinted in P.H. Winston, ed. 1975. The Psychology of Computer Vision: 211-277, McGraw Hill).
18. Nouioua, F. 2003. Extraction de connaissances sur les normes dans un corpus textuel. DEA report. Laboratoire d'Informatique, Univ. Paris-Nord. Sept.
19. Pearl, J. 1996. The Art and Science of Cause and Effect. Lecture delivered at UCLA and printed in Pearl, J. 2000. Causality. Cambridge University Press.
20. Reiter, R. 1980. A Logic for Default Reasoning, in (AI 1980): 81-132.
21. Rumelhart, D.E.; McClelland, J.L. eds. 1986. Parallel Distributed Processing (2 vols.) M.I.T. Press.
22. Schank; R.; Abelson, R.P. 1977. Scripts, Plans, Goals and Understanding Lawrence Erlbaum Ass.
23. tal 1994. Special issue "Approches sémantiques". Traitement automatique des langues 35(1).